



Eur päisches Patentamt  
Europ an Patent Offic  
Office européen des brevets



(11) Publication number : 0 686 931 A2

(12)

## EUROPEAN PATENT APPLICATION

(21) Application number : 95303792.6

(51) Int. Cl.<sup>6</sup> : G06K 9/22, G06F 3/033

(22) Date of filing : 02.06.95

(30) Priority : 06.06.94 US 254382

(43) Date of publication of application :  
13.12.95 Bulletin 95/50

(84) Designated Contracting States :  
DE FR GB

(71) Applicant : XEROX CORPORATION  
Xerox Square  
Rochester New York 14644 (US)

(72) Inventor : Poon, Alex D.  
575 South Rengstorff Avenue, Apt. No. 21  
Mountain View CA 94049 (US)  
Inventor : Cass, Todd A.  
784 Dolores Street  
San Francisco CA 94110 (US)  
Inventor : Weber, Karon Anne  
1330 Union Street, No. 22  
San Francisco CA 94109 (US)

(74) Representative : Johnson, Reginald George et al  
Rank Xerox Ltd  
Patent Department  
Parkway  
Marlow Buckinghamshire SL7 1YL (GB)

(54) Searching and matching unrecognized handwriting

(57) A method and system are disclosed for searching and matching gesture-based data such as handwriting without performing a recognition process on the handwritten gesture data to convert it to a standard computer-coded form. Target data (16) collected as sample data points (20,22) of spatial coordinates over time are concatenated into a single target gesture sequence of sample data points. The sample data points comprising the gesture-based data structure to be searched (the corpus) are grouped into corpus gesture sequences for matching against the target gesture sequence. Matching may be done by any suitable method, and a novel signal comparison technique based on dynamic time warping concepts is illustrated. The result of the matching is a list of the locations of the matching corpus gesture sequences in the corpus, which in turn may be used for further processing, such as the display of an image of the matching corpus gestures for a system user. The ability to determine the existence and location of a gesture in the corpus that matches a target gesture (10) is the basis for performing a variety of additional functions, such as a "find and replace" function and the ability to use gestures as keywords to index a gesture-based data structure without performing recognition on either the keyword gestures or the gesture-based data structure. The technique is suitable for inclusion in any system that accepts gesture-based data, such as a personal digital assistant (PDA) or other pen-based computing device.

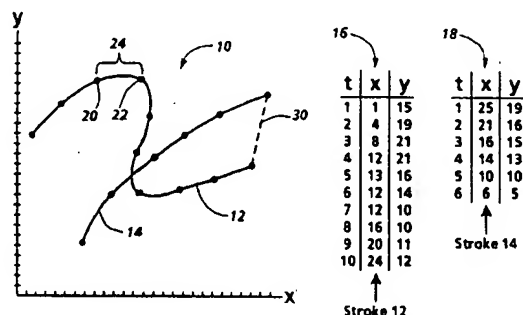


Fig. 1

The present invention relates generally to information processing in systems that accept gestures as input, and more particularly to a system and method for mapping a target gesture-based data item to one or more substantially identical data items in a data structure of data based on gestures, without performing conventional recognition techniques on the gesture-based data.

Some processor-controlled information systems, such as, for example, portable computing devices known variously as notebook computers, personal digital assistants (PDAs), personal communicators, or personal interactive devices, make use of a pen-like input device to enter data in the form of gestures that resemble conventional block letter or cursive handwriting characters and in the form of graphical pictures, shapes or symbols. In some programs, the gesture data are stored as entered. Typically, however, the gesture-based data alone is in a data form that is thought to be of limited or no present utility, and so is discarded once converted to a conventional data form that is of utility to a processor. Thus, many gesture-based data handling systems assume that it is necessary to perform recognition processes on the data based on gestures in order to convert the handwritten or picture input into "recognized," predefined shapes, such as circles, rectangles or straight lines, or into a standard coded representation of characters, such as the standard data format known as ASCII, in order for the data to be suitable for use in standard processor-based applications such as word processing, electronic mail, drawing, graphics, and data base applications. An example of such a system is the Newton® MessagePad available from Apple® Computer, Cupertino, California, which performs a handwriting recognition process on the user's handwritten input and stores the recognized data for use, for example, as calendar entries, addresses and phone numbers, and as text for use in electronic mail messages.

Because of the current problems with handwriting recognition, the utility of systems that accept gestures as input would be improved if some functionality were available without explicitly requiring recognition of the data based on gestures. A software product known as "aha!™ InkWriter™" from aha!™ software corporation, Mountain View, California, provides editing functions that are apparently based on the gesture-based data itself. The aha!™ InkWriter™ Handbook, 1993, indicates that the software provides some functional gesture-based data processing ability to the user, similar to that provided in a conventional word processing application, without first requiring the user to explicitly invoke a handwriting recognition process. The Handbook also indicates that handwritten input may be "translated" into what is referred to as "computer text", or just "text." Illustrations in the Handbook indicate that "translation" corresponds to a

conventional handwriting recognition process.

InkWriter also provides a word or character search capability. Searching involves the user selecting the document that is to be searched, and choosing the Find function from the Edit menu; in response to the selection of the Find function, a Find dialog sheet opens. The user prints the word he or she wants to find. The Handbook states that the word the user prints in the Find dialog sheet is translated immediately. If the word is found, the word "Found" appears in the result section of the Find sheet and the word is selected. Because the search capability described in the aha! InkWriter Handbook depends on performing handwriting recognition on the searched-for word entered by the user, it would be reasonable to conclude that handwriting recognition is performed on the writing in the document being searched, even though the user has not taken an affirmative step to translate the writing into computer text, and despite the fact that the writing does not appear displayed to the user as computer text. The ability to successfully locate words in documents using the inkWriter software, then, depends on the accuracy and robustness of the translation, or recognition, process. The Handbook supports this conclusion by advising the user to print, rather than write in longhand, if the user plans to be doing a lot of searching.

A recognition system converts a gesture-based data input target, such as handwriting, into a computer-coded or human-understandable form by mapping it to, or classifying it as, one of many known templates or prototypes of the gesture that are stored in what may be called a reference data structure; a successful mapping identifies the computer-coded or human-understandable form of the gesture, or identifies the writer of the gesture. Such systems typically fall into one of the categories of handwriting recognition, signature verification, and writer identification.

An important characteristic of each of these types of recognition systems is the nature of the reference data structure to which target inputs are to be mapped. A reference data structure of template or prototype data objects contains what may be called "enrolled" data objects, that is, data objects that are entered into the reference data structure for the specific purpose of enabling the classification of target inputs; the data that comprises each enrolled data object is extensively prepared for the features, characteristics, or variations in the data that may be expected from potential targets. The accuracy and reliability of a recognition system, then, is directly related to the quality of preparation and characterization of the gesture-based data of the reference templates to which a target gesture is mapped. The large variability in handwriting styles among the potential users of PDA devices, along with a variety of environmental factors that may affect or alter a user's normal handwriting, make the development of broad but accurate refer-

ence templates a very difficult task.

As long as handwriting recognition systems produce unreliable, error-prone results, the wider use and utility of systems that accept gestures as input will be significantly limited by the dependence of these systems on applying a recognition process to the gesture-based data.

The present invention provides a novel technique for determining whether a target gesture-based data item occurs in a body of data based on gestures (hereafter, called the "gesture-based data structure"), such as handwritten notes or documents that include cursive or block printed handwriting as well as graphical pictures or symbols, without applying a recognition process to either the target gesture-based data or the gesture-based data structure. The present invention is thus advantageous to the overall usefulness of systems that accept gestures as input by reducing the dependence of these systems on the success of a handwriting recognition process. Determining the occurrence of a target gesture in a gesture-based data structure provides the capability to perform a variety of processing functions, including a search function that may provide a user with an image of the gesture-based data structure highlighting the location(s) of the target gesture, and a function analogous to a "find and replace" function in a word processing application, where the located target gesture may be replaced with a second target gesture. In addition, in conjunction with software that builds a keyword index of a data structure, gestures may be used as keywords to index a gesture-based data structure without performing recognition on either the keyword gestures or the gesture-based data structure.

Another advantage of the matching technique of the present invention is its ability to minimize the impact of the stylistic variations among writers on the accuracy of the matching process by allowing the target gesture used for searching to be selected from the gesture-based data structure that is to be searched, if it exists there. A first writer reviewing the handwritten notes of a second writer that expresses a certain concept or idea could search for all of the occurrences of that concept or idea in the notes by selecting it as a target gesture for searching. The target gesture would then have the stylistic characteristics of the original, second writer and the gestures in the rest of notes, which is likely to improve matching results.

The present invention is premised on the discovery that reasonably accurate gesture matching between a target gesture and a gesture in the gesture-based data structure can be achieved by using any one of several signal comparison techniques to match a target gesture to each gesture in the gesture-based data structure until all gestures have been compared to the target, to determine if and where the target gesture is included in the gesture-based data structure. Because of the variability between writers in entering

gesture-based data, a single gesture provided as the target may occur in the gesture-based data structure as a sequence of two or more gestures, or a single gesture included in the gesture-based data structure may be entered as a sequence of two or more target gestures. The invention provides for concatenating a sequence of target gestures into a single group of sample data points for purposes of performing signal comparison. Then, since it is possible that the target may be matched in the gesture-based data structure by either a single gesture or a sequence of gestures, the invention provides to the signal comparison process, as input with the target gesture, a grouping of sample data points from the gesture-based data structure. Providing such groupings of sample data points from the gesture-based data structure for comparison reduces the impact of the handwriting variations among writers on the invention's ability to make matches, and thus improves accuracy.

Therefore, in accordance with the present invention, there is provided a method of operating a system that includes a signal source for providing data based on gestures, memory for storing data based on gestures, and a processor connected for receiving signals. The method comprises receiving target gesture data from the signal source. The target gesture data is based on a first gesture and includes a first plurality of sample data points indicating the first gesture's path of motion over an interval of time. A gesture-based data structure stored in the system memory is provided that indicates information about gestures and includes, for each respective gesture, a plurality of sample data points indicating a respective path of motion over a respective interval of time of the respective gesture. The second plurality of sample data points are grouped into a plurality of corpus gesture sequences, each including at least two sample data points. Then, the target gesture data is compared to each of the corpus gesture sequences to determine if the target gesture data matches one of the corpus gesture sequences. A list of corpus gesture sequences that match the target gesture data is produced. The list includes, for each corpus gesture sequence, the location in the gesture-based data structure of the matching corpus gesture sequence.

In accordance with another aspect of the present invention, a processor-controlled system is provided that comprises a signal source for providing target gesture-based data including a first plurality of sample data points indicating a first path of motion over a first interval of time of a first gesture; a processor connected for receiving the target gesture-based data from the signal source; and memory for storing data. The data stored in the memory includes instruction data indicating instructions the processor can execute, and a gesture-based data structure that indicates information about gestures and includes, for each respective gesture, a plurality of sample data

points indicating a respective path of motion over a respective interval of time of the respective gesture. The processor is also connected for accessing the data stored in the memory. The processor, in executing the instructions in response to receiving the target gesture-based data from the signal source, groups the second plurality of sample data points into a plurality of corpus gesture sequences, each corpus gesture sequence including at least two sample data points. The processor then compares the target gesture-based data to each of the corpus gesture sequences to determine whether the target gesture-based data matches one of the corpus gesture sequences, and produces a matching list of corpus gesture sequences that match the target gesture sequence. The matching list includes, for each corpus gesture sequence, the location in the gesture-based data structure of the matching corpus gesture sequence.

The present invention will be described further, by way of examples, with reference to the accompanying drawings, in which:-

FIG. 1 illustrates a gesture and data based on a gesture;

FIG. 2 is a simplified block diagram illustrating a system the present invention may operate and illustrating a software product in which the present invention may be implemented;

FIG. 3 is a flow chart illustrating the general steps of searching and matching gesture-based data according to the present invention;

FIG. 4 is a flow chart illustrating the steps of matching a target gesture group to gesture-based data groups in a gesture-based data structure according to a software implementation of the present invention;

FIG. 5 is a flow chart illustrating the substeps of the signal comparison step illustrated in FIG. 4, according to the illustrated implementation of the present invention; and

FIG. 6 is a graph of an exemplary stroke function, constructed for the gesture of FIG. 1, for use in the signal comparison process of FIG. 5, in accordance with this invention.

The following terms provide the framework for describing the embodiment of the claimed invention illustrated in the accompanying drawings.

A "gesture" is an expressive movement. FIG. 1 illustrates gesture 10, which resembles the English letter "x" or which could be an arbitrary graphical symbol. Data or a signal is "based on" a gesture when the data or signal includes information about the gesture's path of motion in space over an interval of time. For example, if information about the gesture's path includes only one point in time, the data or signal is not based on the gesture, but if information about one or more points of the path in space and the time interval of the point or points is included, the data or signal is based on the gesture. A simple example of data

based on a gesture is data indicating the beginning and ending points in space of the gesture, and the time each occurred. More complete data based on a 2-dimensional gesture would be a vector of (x, y, t) 3-tuples obtained by sampling the x, y coordinates of the path of the gesture at a sufficiently high frequency to capture substantially all movement content provided by the gesture. An example of such data is shown in table 16 in FIG. 1. Data based on a gesture may include information about a gesture from which all or a portion of information related to the gesture's path of motion over time may be sensed, computed or derived, such as, for example, a gesture's acceleration or velocity. Additional component information about a gesture may also be captured, such as, for example, data representing the pressure, disturbances in resistance, or acoustic surface waves of the input device producing the gesture on a data collecting surface. Data or a signal based on a gesture is also referred to herein as "gesture-based data." A "gesture-based data structure" includes data or signals based on one or more gestures.

A "stylus" as used herein refers to a user input device capable of producing gesture-based data. For example, an electronic pen-like device is a stylus, and the movement of the tip of the stylus across a data collecting surface produces gesture-based data. A stylus may also include a conventional pointing device, such as a mouse, that is capable of producing data based on a gesture. In that case, the image of the cursor in the display area of the display device represents the location of the stylus, and the display area represents the data collecting surface. The term "electronic ink" or "digital ink" refers to the image on a display of the trace of the motion of a stylus tip, and is generally used to refer to an image of gesture-based data.

A "stroke" is the data or signal from a gesture or a portion of a gesture produced from the point in time the stylus is determined to be producing gesture-based data to a subsequent point in time when the stylus is determined not to be producing gesture-based data. For example, when a stylus produces gesture-based data by movement across a data collecting surface, a stroke is the data produced from the point in time the stylus is sensed to be in contact with the data collecting surface to the subsequent point in time the stylus is sensed not to be in contact with the data collecting surface. In FIG. 1, gesture 10 is shown as being comprised of two strokes 12 and 14.

Data are collected during the production of a stroke at regular time intervals, as illustrated by sample data points 20 and 22 in FIG. 1. A "stroke segment" or "segment" refers to the portion of a stroke between two time sample data points, as illustrated by segment 24. When the final (in time) sample data point of a first stroke 12 is connected to the first sample data point of a second stroke 14, the strokes or

sample points are said to be "concatenated." Concatenation of sequences of sample points may result in the grouping of multiple strokes into a single stroke group. The effect of concatenating sample data points is to introduce a segment 30 into the sequence of sample data points.

A data structure including gesture-based data may also be referred to herein as an "unclassified" data structure, and is distinguishable from a "classified" data structure which includes enrolled data objects such as templates or prototypes used for the purpose of classifying target inputs.

The present invention for searching and matching gesture-based data operates a processor-controlled system having the common components, characteristics, and configuration of system 100 illustrated in FIG. 2. System 100 includes input circuitry 152 for receiving signals from a signal source 154. Signal source 154 may include any signal producing source that produces signals of the type needed by the present invention. Such sources include input devices controllable by a human user that produce signals in response to actions by the user, such as a stylus device 156 that the user moves over the data collecting surface and display area 180 of device 170 when making gestures. Device 170 is commonly called an "electronic tablet," but as noted earlier, device 170 could also be a conventional display device capable of presenting images, or a stylus device 156 alone that is capable of producing gesture-based data without using a data collecting surface could be substituted for the combination of device 170 and stylus device 156. Alternatively, signal source 154 may be an operation (not shown) that processor 140 is executing that provides data based on gestures as a target gesture to processor 140 for processing according to the present invention. Signal source 154 may also be an image scanning device 158 which produces signals defining an image of gestures which are sent through input circuitry 152. Processor 140 may then convert the image of gestures to data based on gestures using any one of a number of known techniques.

Processor 140 operates by accessing program memory 114 of memory 110 to retrieve instructions, which it then executes. Program memory 114 includes gesture data comparison instructions 116 that implement the invention described in the flowcharts of FIGS. 3, 4 and 5. During execution of instructions, processor 140 may access data memory 122 in addition to receiving input signals from input circuitry 152, and providing data defining images, for example, images of electronic ink to output circuitry (not shown) for presentation on device 170 in display area 180, or for presentation on any other suitable display device. In the description of the illustrated embodiment, the display area corresponds to the visible part of the display screen, and the method of the present invention provides for visibly displaying electronic ink therein.

Memory 110 also stores gesture-based data structure 126, target gesture data 130, and, in the case of the illustrated embodiment, stroke function data 132, as well as other data.

The actual manner in which the physical components of machine 100 are connected may vary, and may include hardwired physical connections between some or all of the components, such as shown by the dotted line wire connection 160 between stylus 156 and device 170, as well as connections over wired or wireless communications facilities, such as through remote or local communications networks and infrared and radio connections. The range of the physical size of machine 100 may vary considerably from a very large device, for example, a large electronic "whiteboard" device for shared collaboration, to much smaller desktop, laptop, and pocket-sized or smaller display devices.

FIG. 3 illustrates the general steps of the present invention. Note that the term "corpus" is used as a shorthand term to mean the gesture-based data structure that is being searched for comparisons to a target gesture. Target gesture data is received from a signal source, in box 208. To reduce the matching errors that are introduced by writer variability, the sequence of sample data points for each stroke of the target gesture data are concatenated, in box 214, into a sequence of consecutive sample data points, referred to hereafter as a "target gesture sequence." The effect of concatenating the sample data points of the target gesture data into a single sequence of connected sample points is to introduce a new segment between strokes.

The gesture-based data structure, or corpus, being searched may be any collection of data based on gestures that is stored in memory 110 (FIG. 2) or otherwise accessible to processor 140. Any data structures of data based on gestures representing handwritten notes, letters, logs, diaries, calendars and other documents may become a corpus to be searched. The corpus may be viewed as simply a consecutive stream of sample data points representing the path of motion over time of many gestures; the data stream may be organized into lists of strokes, but need not be. In order to permit the matching of a target gesture sequence to the gesture-based corpus data, sets of sample data points in the corpus are identified and grouped into sequences of sample data points referred to hereafter as "corpus gesture sequences," in box 220. Any one corpus gesture sequence of sample data points may overlap with, or share sample data points in common with, another corpus gesture sequence of sample data points. A corpus gesture sequence is any subset of sample data points in the corpus and may be as small as two sample data points. Each corpus gesture sequence is concatenated, in box 224, into a single data unit, or signal, for comparison purposes.

In box 228, the target gesture sequence is compared to each corpus gesture sequence to determine whether there is a match. Any suitable signal matching process may be used for this purpose; an example of a novel signal matching process based on dynamic time warping techniques is presented in detail below. The results of the matching process of the target gesture sequence against all corpus gesture sequences are then produced, in box 234, and comprise, minimally, a list indicating the corpus gesture sequences that match the target gesture sequence and including, for example, the indices of the starting and ending sample data points of each matching corpus gesture sequence in the corpus, ( $i_{start}$ ,  $i_{end}$ ). The list may be used for a variety of further purposes, including but not limited to simultaneous or later reporting to a human user, typically by displaying the matching results in an image, or to another system operation, or passing a list of matching corpus gestures, or pointers to matching corpus gestures, for further processing by a subsequent operation. The list of matches may be empty, if no match occurs.

FIG. 4 illustrates the sequence of steps implemented by an implemented embodiment of the present invention. The gesture-based data in both the target and corpus is organized into a list of strokes, each stroke being a list of sample data points, as illustrated in FIG. 1. In box 308, data indicating a target gesture is received from a signal source. In box 310, the target gesture data is examined to determine if it includes more than one stroke, and, if it does, each stroke<sub>i</sub> is concatenated with a subsequent stroke<sub>i+1</sub> until all strokes have been concatenated into a single sequence of consecutive sample data points by connecting the last sample point of stroke<sub>i</sub> (the stroke first in time) with the first sample point of the stroke<sub>i+1</sub> (the stroke next in time).

In box 314, a counter,  $n$ , for counting the strokes in the corpus is initialized to one. When all strokes in the corpus have been compared to the concatenated target stroke data, an inquiry that is made in box 374, processing for this target is complete. In box 316, a grouping counter  $m$  is initialized to one; grouping counter  $m$  controls the grouping of the strokes in the corpus for input into the matching process. Starting from each single stroke in the corpus, as counted by counter  $n$ , and beginning with grouping counter  $m$  at one, corpus strokes  $n$  to  $n + m - 1$  are grouped into a single sequence of consecutive sample data points and concatenated, in box 320, before being input into signal comparison process 330, until  $m$  has reached a maximum number of grouping strokes, which is tested in box 368. The maximum value for grouping counter  $m$  may be determined heuristically after experimenting with various types of corpora produced by the same or by different writers. Alternatively, the maximum value for grouping counter  $m$  may be a function of the number of strokes,  $s$ , in the target ges-

ture data. In the illustrated implementation, grouping counter  $m$  is a function of  $s$ , and its maximum value is equal to  $s + 2$ .

This implementation of the concept of grouping leads to a brute force signal comparison of groups of corpus strokes to the concatenated target gesture data that attempts to account for variations among writers that result from inconsistent and different stroke spacing. More importantly, this approach does not depend on the ability of the system to divide the corpus into discrete semantic units, such as word or graphical symbols, in advance of searching and matching. Such a brute force comparison method, however will have a substantial impact on computational efficiency. Grouping counter  $m$  is a processing and performance efficiency control. In theory, given a corpus the size of  $k$  strokes, at a point  $n$  in the sequence, the value of  $m$  could range from 1 to  $k - n + 1$ , and strokes would then be grouped into sequences as large as the entire corpus to ensure that the maximum number of stroke sequences are input to the signal comparison process. In the illustrated implementation, a target gesture with a stroke count,  $s = 3$  would result in 5 corpus gesture sequence group, ranging from 1 stroke to 5 consecutive strokes (when the maximum value of  $m = s + 2$ ), each group being compared to the target gesture, for each stroke in the corpus.

It can be seen that a variety of different approaches can be used to determine grouping by altering either or both of the initial value for group counter  $m$  and its maximum value. For example, group counter  $m$ , could be initialized to  $s - 1$ , and its maximum value could be established at  $s + 1$ , so that only groupings one stroke less than the target, equal to the target, and one stroke greater than the target are formed for comparison with the target. Grouping counter  $m$  has to be selected carefully enough to ensure that a target gesture with a small number of strokes will be successfully matched with corpus gesture sequences having a larger number of strokes, which might occur when a target composed of substantially cursive handwriting needs to be matched to a corpus composed of substantially printed handwriting.

Signal comparison process 330, described in more detail below, returns a score that measures how well the concatenated target gesture data compared with a particular group of concatenated strokes. This score is evaluated against a threshold score, in box 360, and if it compares favorably - in this case, if the score is less than the threshold - the group of concatenated corpus strokes is marked as a match. Marking in this sense simply refers to any system action taken in response to finding a matching corpus gesture sequence. In the illustrated embodiment, the corpus being searched is displayed for a user on an electronic tablet. As each match of the concatenated target gesture sequence is encountered in the corpus, a box-



shaped display feature is added around the portion of the image of the corpus representing the matching corpus gesture sequence, and when the next render-image instruction (e.g., paint-image or display-image instruction) is executed, the system presents an image of the corpus that includes the matching corpus gesture sequences with boxes surrounding them.

If the score resulting from signal comparison process 330 does not compare favorably to the threshold, and after a successful match is marked, grouping counter *m* is incremented, in box 366, in order to form the next group of corpus strokes. If grouping counter *m* is not greater than its maximum, the next group of strokes in the corpus is identified and concatenated, in box 320, in preparation for signal comparison with the target gesture group.

When grouping counter *m* is greater than its maximum, corpus stroke counter *n* is incremented, in box 370, tested against the number of strokes in the corpus to determine if all corpus strokes have been compared, and, if not, grouping counter *m* is reset to one in box 316 in order to begin a new iteration of grouping with the next consecutive stroke in the corpus.

The threshold represents the cutoff in the score of a signal comparison that determines that the target gesture sequence and a corpus gesture sequence match sufficiently enough to report as a match, and is, to a large extent, a function of the design decisions made with respect to the implementation of signal comparison process 330, such as, for example, how normalization is done. In contrast to recognition systems that use a lowest score to indicate a successful comparison, the threshold is somewhat a subjective and arbitrary assessment of how many false positives will be tolerated by the system or system user. In this illustrated implementation, the value of the threshold has been determined empirically, and a threshold of 4 has been used that results in substantially reducing false positives without unduly increasing false rejections.

In the illustrated implementation, signal comparison process 330 uses a mathematical comparison technique known as dynamic time warping (DTW), which refers to the comparison of trajectories, which is defined to be a continuous function of time in multidimensional space. The basic idea of time warping is that replications of the same trajectory will trace out approximately the same curve, but with varying time patterns. Time warping techniques compare sequences derived from trajectories by time sampling, when each trajectory is subject not only to alteration by the usual additive random error but also to variations in speed from one portion to another. Such variation in speed appears concretely as compression and expansion with respect to the time axis, and is referred to as compression-expansion. Time warping deals with such variation, and can be used to measure how different two sequences are in a way that is

not sensitive to compression-expansion but is sensitive to other differences.

The present invention applies the concepts of time warping to the domain of the gesture or stroke, which can be said to be a trajectory. The variation among writers that is of major interest is the variation that results from differences in stroke shapes; these variations are captured in the angles and distances between sample data points, and, in this domain, the "stroke trajectory" of a sequence of sample points is actually a function, hereafter called a "stroke function," that relates the orientation of a segment to the cumulative length of previous segments. In the literature this type of representation is often called the "orientation" function or the "orientation versus arc length" function. An example of a stroke function is shown in FIG. 6, in which graph 40 shows the stroke function for the concatenated strokes of gesture 10 in FIG. 1. In graph 40, the cumulative length at each stroke segment of concatenated strokes 12 and 14 (FIG. 1) along the horizontal axis 42 of graph 40 is plotted against the segment's angle on the vertical axis 44 of graph 40. The stroke function in graph 40 is an example of the target and corpus stroke functions that are input to the DTW comparison process.

In the mathematical sense, the idea of time warping is to establish a correspondence between each point in the target stroke function and some specific point in the corpus stroke function, by warping the time axis of each stroke function and establishing a mapping between points in the two stroke functions. Two time warpings are equivalent if they induce the same linking throughout the entire two stroke functions. The distance, *D*, between two stroke functions is the normalized integral, or summation in the discrete case, of weighted pointwise distances, *d*, between points brought into correspondence by the warping which minimizes this integral or sum. The distance, *D*, serves as the matching score, or measure of the similarity, for a target gesture group and the current corpus gesture group. The normalization referred to is based on the length of the warped sequences. An example of a pointwise distance metric  $d(x,y)$  is  $d(x,y) = |x-y|$ .

FIG. 5 illustrates the detail substeps of the DTW signal comparison process in box 330 of FIG. 4. The stroke functions for the target gesture group and the current corpus gesture group are constructed, in boxes 334 and 338, by dividing the target gesture group and the current corpus gesture group into segments and then calculating the length and angle of each segment. For efficiency, the previous segment list for the target gesture group and for the previous corpus gesture group are stored and reused where possible. The list of segments, angles and lengths for each of the target gesture and corpus gesture groups will include those for the segment that was added as a result of the concatenation in boxes 310 and 320

(FIG. 4.) Use of the dynamic time warping technique minimizes the impact on accuracy of adding a segment to a stroke group as a result of concatenation, since the correspondence of segments between stroke functions can be adjusted to de-emphasize the added segment when its length and angle differ substantially from that of a corresponding segment.

Time warping is then performed on the stroke functions, in box 340, and the distance score is computed and normalized. The score is then returned for testing against the threshold in box 360 of FIG. 4.

Several additional steps may be added to the method described in the illustrated implementation of FIG. 4 in order to improve the matching accuracy or to optimize the performance of the method. One such step involves adding a stroke preprocessing step to the preparation of the respective target and corpus strokes for signal comparison. A stroke preprocessing step, which is, in effect, a noise removal step, would detect and remove stroke or segment irregularities that are likely to result from environmental factors, such as deficient circuitry in the electronic tablet or stylus, rather than from intentional movement by the user. Such a preprocessing step may be accomplished using known techniques.

Another technique that may be used is that of resampling the stroke data points to reduce the number of sample points. In a resampling step, redundant sample points would simply be removed, or the original sample points would be replaced with a new set of sample points that approximates the original curve and removes the redundant sample points that cause excessive computation time. Resampling techniques are also known in the art.

The present invention also contemplates using a scanned image of handwriting as the original source of the corpus, or for specifying a target, or both. Known methods exist for converting a bit-mapped image of handwriting symbols into data based on a gesture.

The present invention may be included as part of the functionality and user interface of any PDA device so that a user of the device would be able to search handwritten corpora.

The system user may select the target gesture to be matched from the corpus to be searched, or from another corpus. An image is displayed of the electronic ink representing all or a portion of the corpus to be used for target selection, and, using the stylus, the user makes a selection gesture indicating the portion of the electronic ink that represents the gesture-based data that are to be used as the target gesture-based data. In response to the selection gesture, the selected gesture-based data is retrieved from the corpus for processing as the target gesture-based data according to the steps illustrated in the flowcharts of FIGS. 3 or 4. This feature may be used by a first writer to search the contents of a corpus written by a second

writer using the second writer's gesture as a target gesture.

Another grouping process that has been implemented groups the corpus into semantically meaningful units. These units are then used as the corpus gesture sequences to which the target gesture-based data are compared during the matching phase. In this implementation, these units are intended to correspond to individual words of a written document. Semantic grouping of the corpus is implemented by first searching through the strokes of the corpus to find horizontal spaces between strokes that are at least half as large as the average vertical size of the strokes in the corpus. Then, semantically-based corpus gesture sequences that could be considered to be roughly equivalent to words, are identified as the sample data points of the strokes that are between the large horizontal spaces.

The order of the concatenated gestures of either the target gesture or those provided from the gesture-based data structure, or both, may be changed prior to performing the signal comparison, to account for other writer variations, for example in the order of crossing letters such as "f" or "t".

Another signal comparison process that may be used is the known technique of simple signal correlation. Correlation also uses a stroke function derived from the sample point sequence formed from the concatenated strokes of the target gesture. The stroke function is correlated with a stroke function derived from the sequence of sample data points in the corpus. Subsequences in the corpus stroke function where this correlation value is high indicate places where there is a good match. In place of pointwise multiplication used in basic correlation, the pointwise absolute difference may be used instead, in a manner similar to that used in the dynamic time warping technique of the illustrated embodiment.

Another signal comparison process that may be used is the technique known as matching based on the Hausdorff distance. Hausdorff matching is a technique for measuring how well two groups of 2D points differ from each other. As applied to the sample data points that comprise gestures, the sample data points of the target gesture, taken as points in the plane, would be compared against groups of sample data points from the corpus, again taken as points in the plane.

The novel technique of the present invention operates on a gesture-based data structure that is wholly unclassified, in the sense that it is a sequence of signals that is undifferentiated as to content, and it is not a data structure of templates or prototypes to which a target is to be specifically classified.



## Claims

1. A method of operating a system including a signal source for providing data based on gestures, memory for storing data based on gestures, and a processor connected for receiving signals; the method comprising:
  - receiving target gesture-based data from the signal source; the target gesture-based data including a plurality of target sample data points indicating a first path of motion over a first interval of time of a target gesture;
  - providing a gesture-based data structure, referred to hereafter as a corpus data structure, stored in the system memory; the corpus data structure indicating information about a plurality of corpus gestures and including, for each respective corpus gesture, a plurality of corpus sample data points indicating a respective path of motion over a respective interval of time of the respective corpus gesture;
  - grouping the respective pluralities of corpus sample data points into a plurality of corpus gesture sequences of sample data points; each corpus gesture sequence including at least two sample data points;
  - comparing the target gesture-based data to each of the corpus gesture sequences to determine whether the target gesture-based data matches at least one of the corpus gesture sequences; and
  - producing a matching list of matching corpus gesture sequences that match the target gesture-based data; the list including, for each matching corpus gesture sequence, a location in the corpus data structure of the matching corpus gesture sequence.
2. A method as claimed in claim 1, wherein the target gesture-based data includes first and second strokes that are consecutive in time; and wherein the method further includes, prior to comparing the target gesture-based data to each of the corpus gesture sequences, concatenating together the last sample data point of the first stroke with the first sample data point of the second stroke into a target gesture sequence of the first plurality of sample data points; each corpus gesture sequence being compared to the target gesture sequence.
3. A method as claimed in claim 1 or claim 2, further including
  - producing an image using the matching list; the image including display features indicating at least one matching corpus gesture sequence included in the matching list; and
  - presenting the image to a system user on
- a display device included in the system.
4. A method as claimed in any one of claims 1 to 3, wherein, when the target gesture-based data does not match any one of the corpus gesture sequences, the matching list is empty.
5. A method as claimed in any one of claims 1 to 4, wherein the signal source is an image scanning device providing image definition data defining an image including a plurality of display features; and wherein the method further includes a converting step for converting the image definition data defining the image including a plurality of display features to the target gesture-based data including the first plurality of sample data points.
6. A method as claimed in claim 5, wherein the display features included in the image represent gestures.
7. A method as claimed in any one of claims 1 to 6, wherein the gesture-based data structure stored in the system memory was converted from image definition data defining an image including a plurality of display features.
8. A method as claimed in claim 7, wherein the display features included in the image represent gestures.
9. A processor-controlled system comprising:
  - a signal source for providing target gesture-based data including a first plurality of sample data points indicating a first path of motion over a first interval of time of a first gesture;
  - a processor connected for receiving the target gesture-based data from the signal source; and
  - memory for storing data; the data stored in the memory including instruction data indicating instructions the processor can execute, and a gesture-based data structure; the gesture-based data structure indicating information about gestures and including, for each respective gesture, a plurality of sample data points indicating a respective path of motion over a respective interval of time of the respective gesture;
  - a processor being further connected for accessing the data stored in the memory;
  - the processor, in executing the instructions in response to receiving the target gesture-based data from the signal source,
    - grouping the second plurality of sample data points into a plurality of corpus gesture sequences; each corpus gesture sequence including at least two sample data points;
    - comparing the target gesture-based data

to each of the corpus gesture sequences to determine whether the target gesture-based data matches one of the corpus gesture sequences; and

producing a matching list of corpus gesture sequences that match the target gesture-based data; the matching list including, for each corpus gesture sequence, the location in the gesture-based data structure of the matching corpus gesture sequence.

10. An article of manufacture for use in a processor-controlled system that includes a user input device for receiving signals indicating actions and requests of a system user; memory for storing data; a storage medium access device for accessing a medium that stores data; and a processor connected for receiving data from the user input device and for accessing the data stored in the memory; the processor further being connected for receiving data from the storage medium access device; the article comprising:

a data storage medium that can be accessed by the storage medium access device when the article is used in the system; and

data stored in the data storage medium so that the storage medium access device can provide the stored data to the processor when the article is used in the system; the stored data comprising

target gesture receiving instruction data indicating input instructions the processor can execute to receive target gesture-based data from the user input device; and

operation performing instruction data indicating response instructions the processor can execute to perform operations for searching for the target gesture-based data in a gesture-based data structure stored in the memory in response to the target gesture-based data;

when the target gesture-based data is received, execution of the response instructions causing the processor to group a plurality of sample data points included in the gesture-based data structure into a plurality of corpus gesture sequences of sample data points, and to compare the target gesture-based data to each of the corpus gesture sequences to determine whether the target gesture-based data matches at least one of the corpus gesture sequences;

when a matching corpus gesture sequence is matched to the target gesture-based data, execution of the response instructions causing the processor to produce a matching list including the location of the matching corpus gesture sequence.

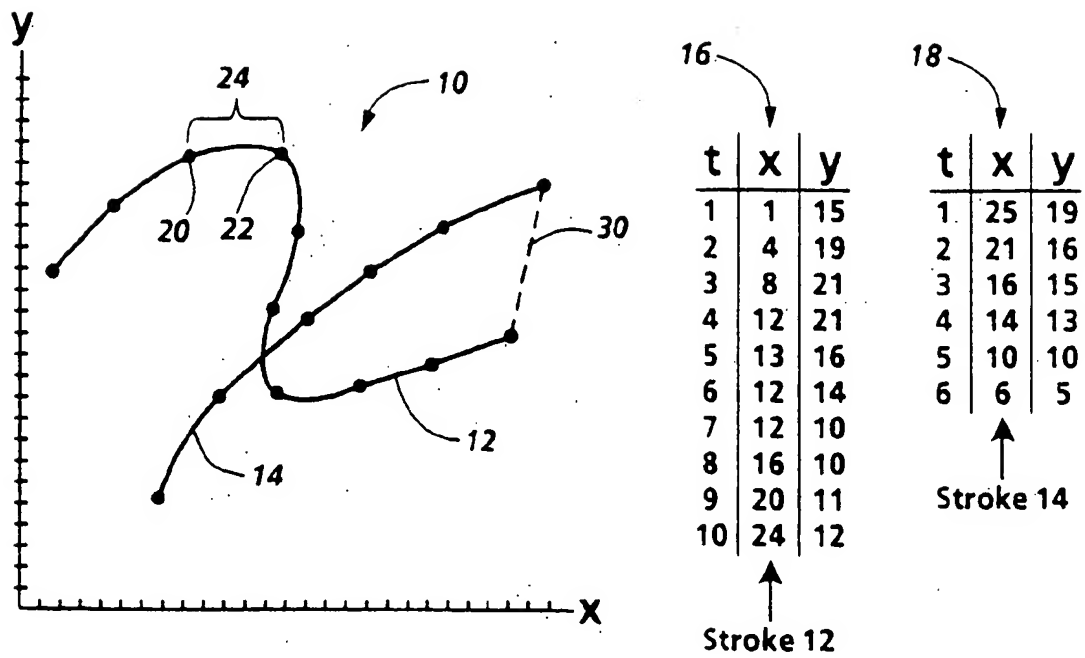


Fig. 1

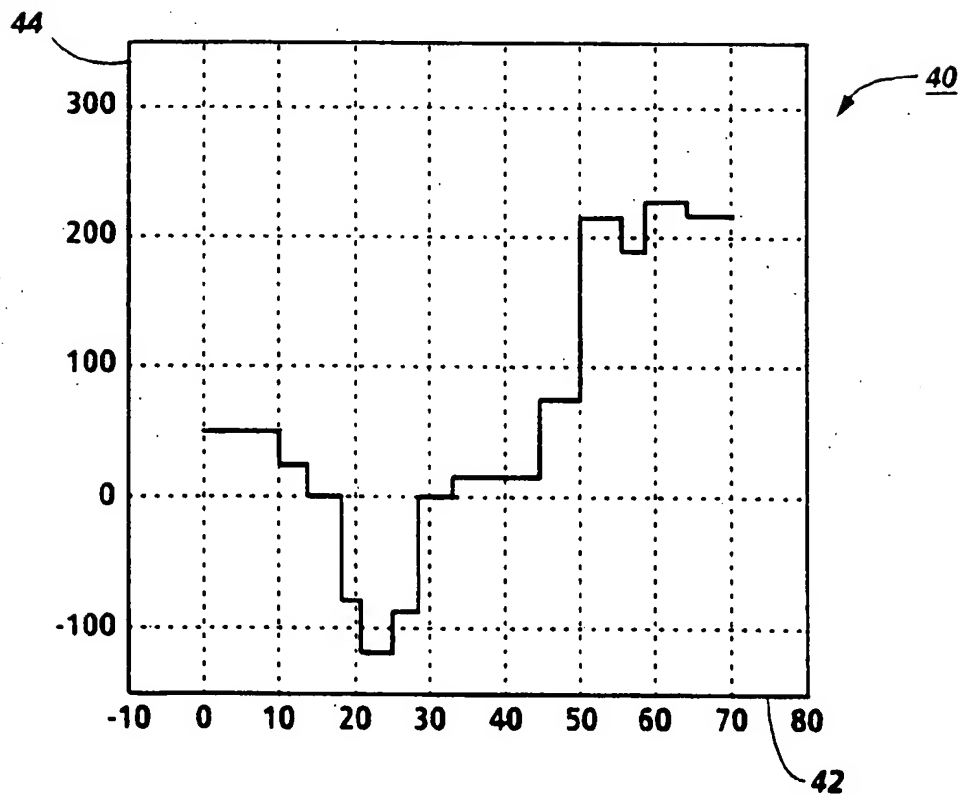
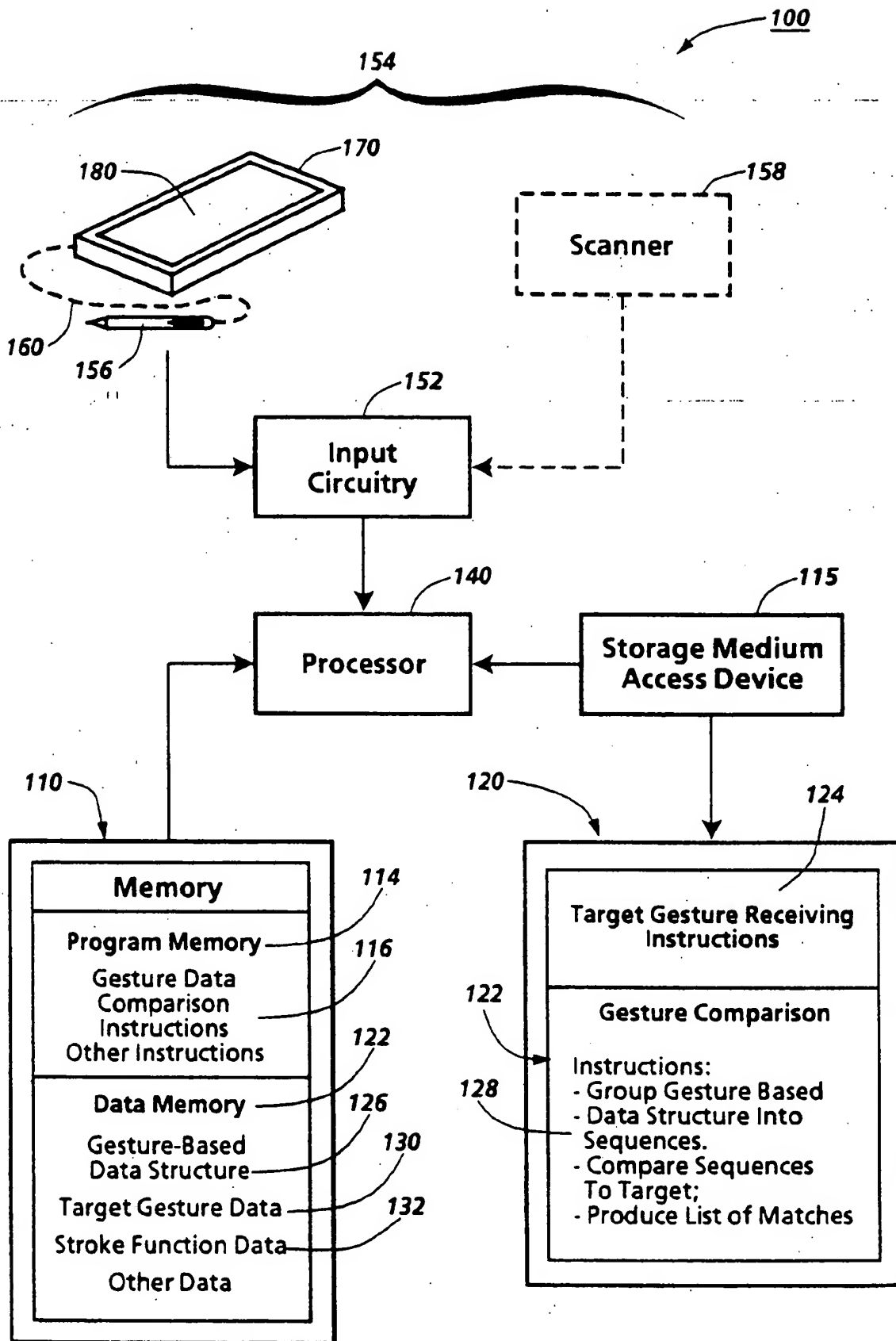
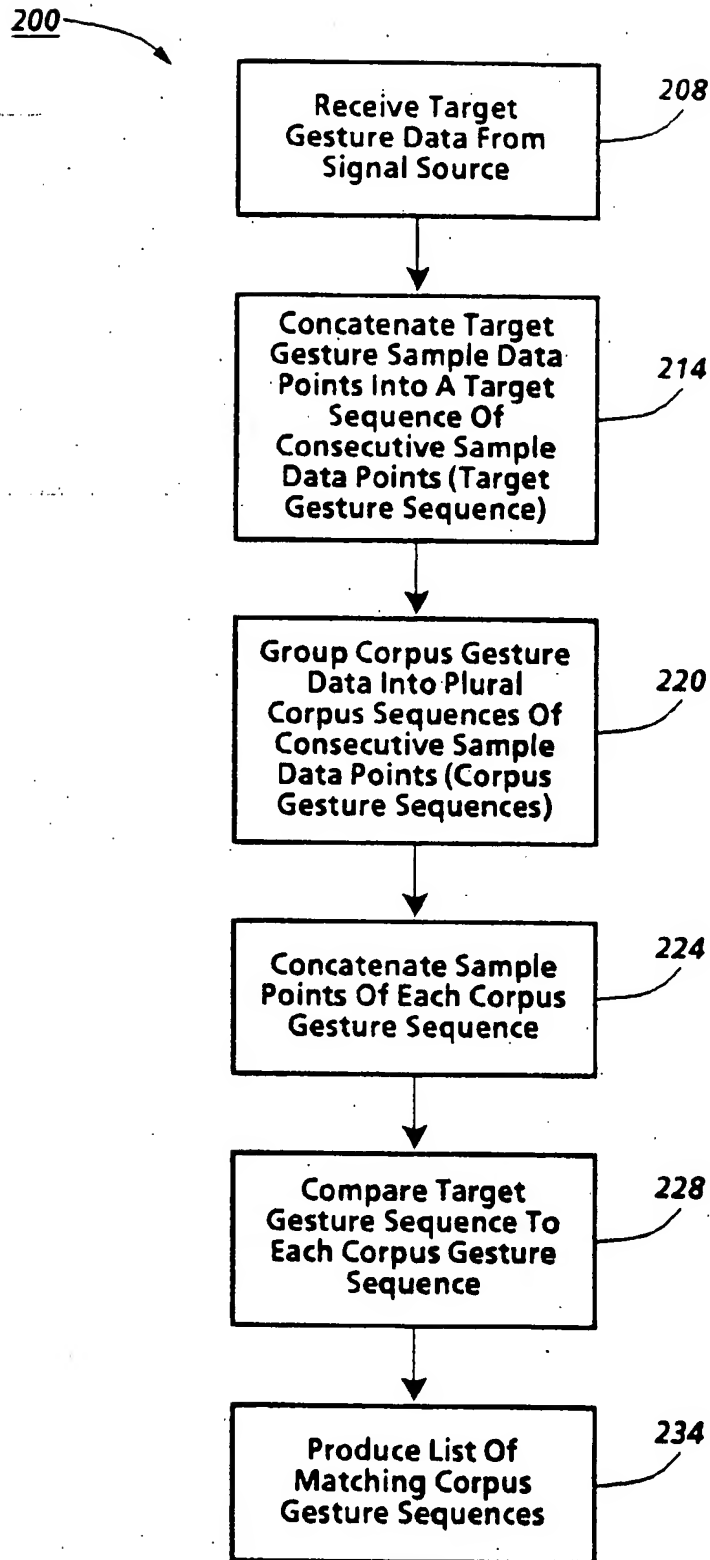
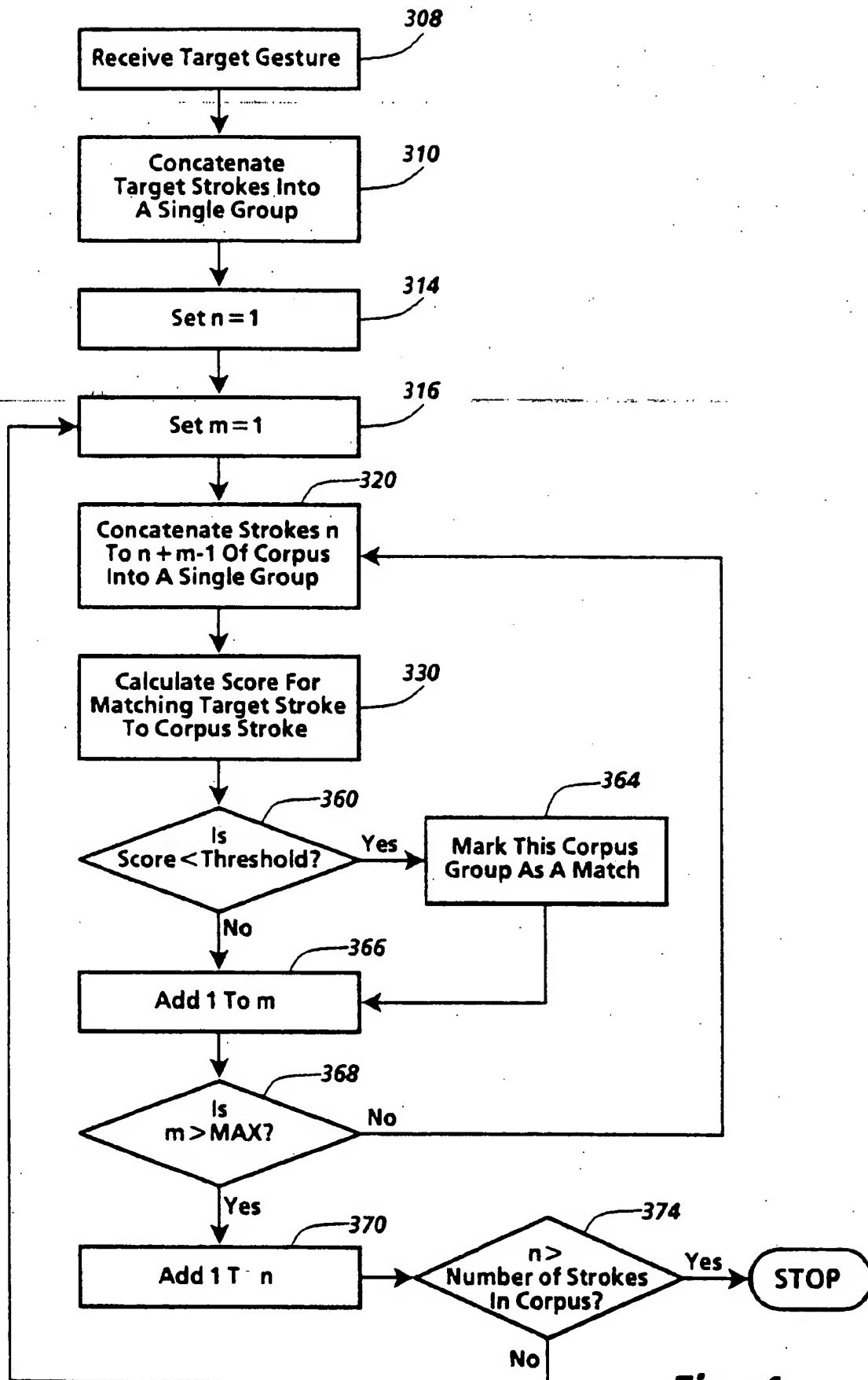


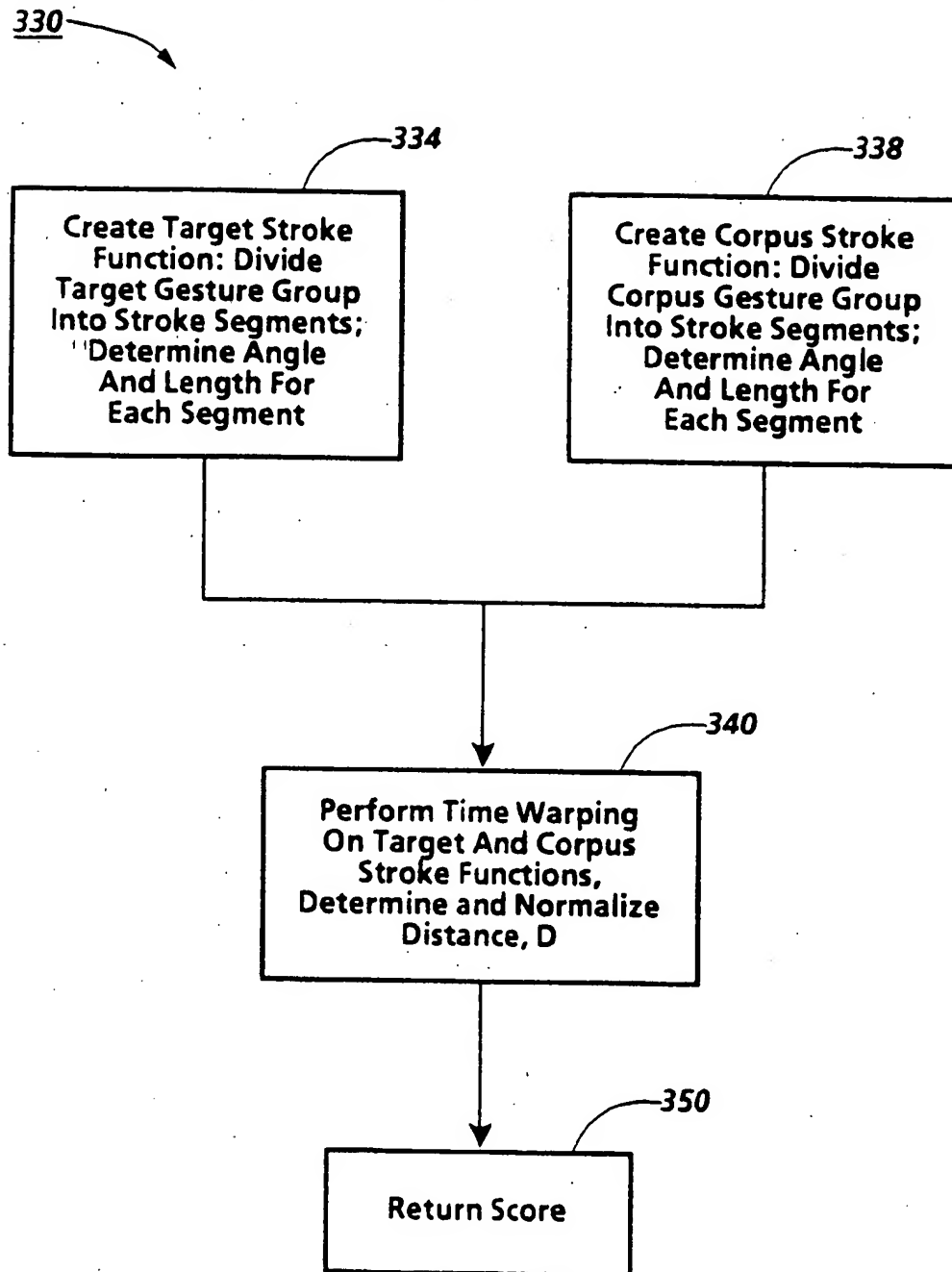
Fig. 6

**Fig. 2**

**Fig. 3**

**Fig. 4**



**Fig. 5**





## EUROPEAN PATENT APPLICATION

(88) Date of publication A3:  
08.01.1997 Bulletin 1997/02

(51) Int Cl.<sup>6</sup>: G06K 9/22, G06F 3/033

(43) Date of publication A2:  
13.12.1995 Bulletin 1995/50

(21) Application number: 95303792.6

(22) Date of filing: 02.06.1995

(84) Designated Contracting States:  
DE FR GB

- Cass, Todd A.  
San Francisco CA 94110 (US)
- Weber, Karon Anne  
San Francisco CA 94109 (US)

(30) Priority: 06.06.1994 US 254382

(71) Applicant: XEROX CORPORATION  
Rochester New York 14644 (US)

(74) Representative: Johnson, Reginald George et al  
Rank Xerox Ltd  
Patent Department  
Parkway  
Marlow Buckinghamshire SL7 1YL (GB)

(72) Inventors:  
• Poon, Alex D.  
Mountain View CA 94049 (US)

## (54) Searching and matching unrecognized handwriting

(57) A method and system are disclosed for searching and matching gesture-based data such as handwriting without performing a recognition process on the handwritten gesture data to convert it to a standard computer-coded form. Target data (16) collected as sample data points (20,22) of spatial coordinates over time are concatenated into a single target gesture sequence of sample data points. The sample data points comprising the gesture-based data structure to be searched (the corpus) are grouped into corpus gesture sequences for matching against the target gesture sequence. Matching may be done by any suitable method, and a novel signal comparison technique based on dynamic time warping concepts is illustrated. The result of the match-

ing is a list of the locations of the matching corpus gesture sequences in the corpus, which in turn may be used for further processing, such as the display of an image of the matching corpus gestures for a system user. The ability to determine the existence and location of a gesture in the corpus that matches a target gesture (10) is the basis for performing a variety of additional functions, such as a "find and replace" function and the ability to use gestures as keywords to index a gesture-based data structure without performing recognition on either the keyword gestures or the gesture-based data structure. The technique is suitable for inclusion in any system that accepts gesture-based data, such as a personal digital assistant (PDA) or other pen-based computing device.

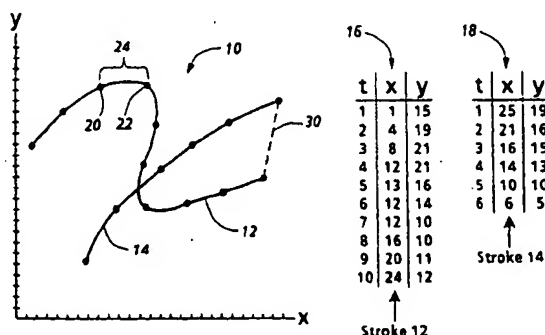


Fig. 1



European Patent  
Office

# EUROPEAN SEARCH REPORT

Application Number  
EP 95 30 3792

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
X	US-A-5 191 622 (SHOJIMA HIROSHI ET AL) 2 March 1993 * abstract *	1-4,7-10	G06K9/22 G06F3/033
A	EP-A-0 464 467 (FROESSL HORST) 8 January 1992 * abstract *	1-10	
A	NEC RESEARCH AND DEVELOPMENT, vol. 34, no. 2, 1 April 1993, pages 263-273, XP000322020 KYOJI HIRATA ET AL: "ROUGH SKETCH-BASED IMAGE INFORMATION RETRIEVAL" * abstract; figure 1 *	1-10	
E	EP-A-0 691 623 (HEWLETT PACKARD CO) 10 January 1996 * abstract *	1,9,10	
			TECHNICAL FIELDS SEARCHED (Int.Cl.6)
			G06K G06F
The present search report has been drawn up for all claims			
Place of search THE HAGUE		Date of completion of the search 8 November 1996	Examiner Sonius, M
<p><b>CATEGORY OF CITED DOCUMENTS</b></p> <p>X : particularly relevant if taken alone  Y : particularly relevant if combined with another document of the same category  A : technological background  O : non-written disclosure  P : intermediate document</p> <p>T : theory or principle underlying the invention  E : earlier patent document, but published on, or after the filing date  D : document cited in the application  L : document cited for other reasons  &amp; : member of the same patent family, corresponding document</p>			

EPO FORM 1500 (12.92) (P04/C01)